第28卷 第3期　　工 程 数 学 学 报　　Vol. 28 No. 3
2011年06月　　CHINESE JOURNAL OF ENGINEERING MATHEMATICS　　June 2011

文章编号:1005-3085(2011)03-0285-08

一种基于密度核估计的最大熵方法

刘　钰[1]，韩　峰[1]，王玉恒[1]，乔登江[1,2]，王建国[1,3]

(1- 西北核技术研究所，西安　710024;
2- 上海华东师范大学，上海　200062;　3- 西安交通大学，西安　710049)

摘　要: 针对未知概率分布类型的测量数据，采用最大熵方法可以确定被测量含有最少人为主观假设的概率分布．但经典最大熵方法由于其密度函数的数学形式导致全局优化难度较大，计算结果难以满足概率密度函数基本性质，需要截断及归一才能使用．本文通过引入概率密度核估计方法，对最大熵方法进行了改进，简化了计算过程，并证明了改进后计算结果能够满足概率密度函数基本性质，最后通过算例说明了该方法的有效性．

关键词: 测量数据；最大熵方法；密度核估计；概率分布估计

分类号: **AMS**(2000) 62F15　　**中图分类号**: O212.1　　**文献标识码**: A

1　引言

最大熵方法 MEM (Maximum Entropy Method) 是 Jaynes 于 1957 年提出的一种可处理不完全信息的方法，利用信息论中熵的概念来尽量避免主观因素对推断的影响[1]．对于只有测量数据的情况，如果没有充足的理由认为其服从某种概率分布，则可通过最大熵方法确定出其最不带倾向性的分布形式及参数[2]，且运用最大熵原理得到的密度函数是被测量的最小偏差估计[3]．

经典最大熵方法利用拉格朗日乘子法导出了概率密度函数估计的解析形式，并构造优化问题．但该估计的解析形式对优化求解产生了一定制约，计算结果难以满足概率密度函数的基本性质，甚至存在不严格可积的现象，这时往往需要截断及归一处理后才能使用．本文提出了一种结合密度函数核估计的最大熵方法，该方法利用密度核估计函数能够收敛于任何复杂密度函数的性质，对上述优化问题进行了重新描述，在熵值最大意义下给出了满足约束条件的被测量密度函数估计，能够避免经典最大熵方法存在的问题．

2　最大熵方法基本原理及问题

最大熵方法是按最大熵原理求解不适定问题的方法[4,5]．具体地，随机变量 x 的概率密度函数 $p(x)$ 的信息熵可定义为

$$H(p)=-\int_R p(x)\ln p(x)\,\mathrm{d}x, \tag{1}$$

收稿日期: 2009-09-24.　**作者简介**: 刘钰 (1982年11月生)，男，硕士，助理研究员．研究方向：武器系统评估技术．

其中 R 为积分空间[1]，则优化问题可以描述为

$$\begin{aligned}\max \quad & H(p) = -\int_R p(x)\ln p(x)\,\mathrm{d}x \\ \text{s.t.} \quad & \int_R p(x)\,\mathrm{d}x = 1, \\ & \int_R x^i p(x)\,\mathrm{d}x = M_i, \quad i = 1, 2, \cdots, m,\end{aligned} \tag{2}$$

其中 M_i 为第 i 阶原点矩，m 为所用最高阶矩的阶数．通过调整 $p(x)$ 使其熵值达到最大值，采用拉格朗日乘子法可以解得

$$\hat{p}(x) = \exp\Big(\lambda_0 + \sum_{i=1}^{m}\lambda_i x^i\Big), \tag{3}$$

上式就是熵值最大的概率密度函数的解析形式．根据 (3) 式并结合 (2) 式中的矩约束条件，可进一步建立求解 $\lambda_1, \lambda_2, \cdots, \lambda_m$ 的 m 个方程组

$$M_i = \frac{\displaystyle\int_R x^i \exp\Big(\sum_{i=1}^{m}\lambda_i x^i\Big)\mathrm{d}x}{\displaystyle\int_R \exp\Big(\sum_{i=1}^{m}\lambda_i x^i\Big)\mathrm{d}x}. \tag{4}$$

在数值求解中，可将 (4) 式中的非线性方程组转化为非线性优化问题，令

$$R_i = 1 - \frac{\displaystyle\int_R x^i \exp\Big(\sum_{i=1}^{m}\lambda_i x^i\Big)\mathrm{d}x}{\displaystyle M_i\int_R \exp\Big(\sum_{i=1}^{m}\lambda_i x^i\Big)\mathrm{d}x}, \tag{5}$$

R_i 为残差，用无约束非线性优化方法可以求出这些残差平方和的最小值．

上述方法可以确定出最大熵密度函数 $\hat{p}(x)$ 的参数 $\lambda_1, \lambda_2, \cdots, \lambda_m$，再将 (3) 代入 (2) 式中的归一化条件，可以解出 λ_0，参见文献 [6]．参数 $\lambda_1, \lambda_2, \cdots, \lambda_m$ 初值的选择对求解的收敛相当重要，初值确定不当，可能会导致求解失败或搜索不到全局最优解[7]，特别是当 m 取值较大时，密度函数 $\hat{p}(x)$ 的数学形式会给求解过程带来不小的困难．另外，在实际应用中，归一化条件往往无法满足，以 3 阶矩最大熵方法计算结果为例，在某些参数条件下该方法得到的密度函数曲线，如图 1 所示．

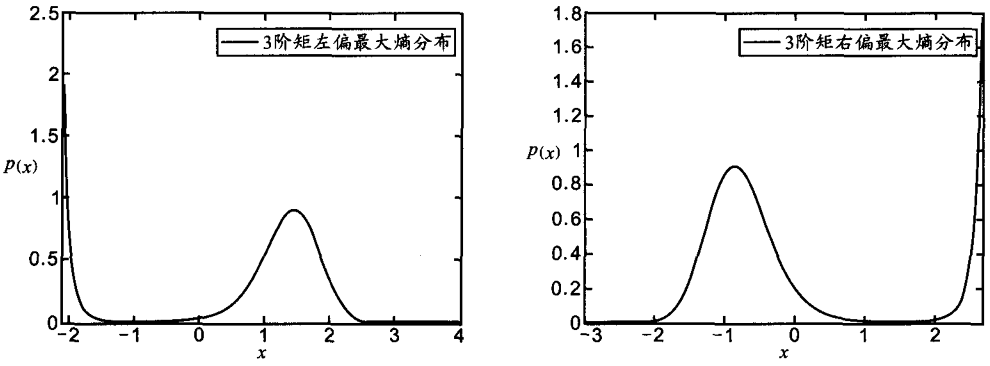
显然，无论被测量数据偏度情况如何，都可能出现密度函数曲线一侧陡增的情况，这样的曲线是不可积的，需要对计算结果截断及归一后才能作为概率密度函数使用，这样，势必增大了估计的偏差．

3 基于密度核估计的最大熵方法

在符合最大熵方法思想的情况下，被测量的概率密度估计 $\hat{p}(x)$，如果能够满足密度函数的一般性质：

1) $p(x) \geq 0$， 2) $\displaystyle\int_{-\infty}^{+\infty} p(x)\mathrm{d}x = 1$，

则可以显著减少截断误差．通过引入密度函数核估计方法，这种设想即可实现．密度核估计方法可以较好地应用于规则分布或不规则分布，单峰或多峰分布的估计，而且只要样本容量足够大，总可以保证以任意精度收敛于任何复杂的未知密度[8]．



(a) 概率密度函数左侧不可积情况　　(b) 概率密度函数右侧不可积情况

图 1: 3阶矩最大熵密度函数

3.1 密度函数核估计基本原理

设 $K(u)$ 为 R 上的一个给定的概率密度函数，则

$$p_n(x)=\sum_{i=1}^{n}\alpha_i\Big[\frac{1}{h_i}K\Big(\frac{x-x_i}{h_i}\Big)\Big] \tag{6}$$

称为总体密度 $p(x)$ 的一个核估计[8]，$K(u)$ 为核函数 (也称为窗函数)，n 为核函数的数量，$h_i>0$ 为窗宽，α_i 是每个核函数的权重[9,10]，且 $0<\alpha_i<1$，$\sum\limits_{i=1}^{n}\alpha_i=1$, $i=1,2,\cdots,n$.

因为核函数 $K(u)$ 为概率密度函数，则满足：

1) $K(u)\geq 0$，　2) $\displaystyle\int_{-\infty}^{+\infty}K(u)\mathrm{d}u=1$，

显然，在 $K(u)\geq 0$ 的限制下，自然能保证 $p_n(x)$ 的非负性，而且可以证明

$$\begin{aligned}\int_R p_n(x)\mathrm{d}x &= \int_R\sum_{i=1}^{n}\alpha_i\Big[\frac{1}{h_i}K\Big(\frac{x-x_i}{h_i}\Big)\Big]\mathrm{d}x\\ &=\sum_{i=1}^{n}\alpha_i\int_R\frac{1}{h_i}K\Big(\frac{x-x_i}{h_i}\Big)\mathrm{d}x=\sum_{i=1}^{n}\alpha_i\int_R K(u)\mathrm{d}u=\sum_{i=1}^{n}\alpha_i=1,\end{aligned} \tag{7}$$

故 $p_n(x)$ 满足归一化条件，即：

1) $p_n(x)\geq 0$，　2) $\displaystyle\int_{-\infty}^{+\infty}p_n(x)\mathrm{d}x=1$，

这就证明了 $p_n(x)$ 是一个合理的概率密度函数，满足了作为总体密度 $p(x)$ 估计的基本条件[8].

3.2 基于密度核估计的最大熵方法

根据最大熵方法的思想，将核估计概率密度函数 $p_n(x)$ 取符合约束条件且信息熵值最大的状态，则

$$H(p_n)=-\int_R\sum_{i=1}^{n}\alpha_i\Big[\frac{1}{h_i}K\Big(\frac{x-x_i}{h_i}\Big)\Big]\cdot\ln\Big(\sum_{i=1}^{n}\alpha_i\Big[\frac{1}{h_i}K\Big(\frac{x-x_i}{h_i}\Big)\Big]\Big)\mathrm{d}x. \tag{8}$$

令 $p_n(x)$ 的信息熵达到最大，则优化问题为

$$\begin{aligned} \max \quad & H(p_n) \\ \text{s.t.} \quad & \int_R x^i p_n(x)\,\mathrm{d}x = M_i, \quad i=1,2,\cdots,m, \end{aligned} \tag{9}$$

其中 M_i 为第 i 阶原点矩的值，m 为最高阶矩的阶数.

本文将这样构造的最大熵方法称为密度核估计最大熵方法KDE-MEM (Kernel Density Estimate-Maximum Entropy Method). 由于概率密度函数 $p_n(x)$ 的性质，保证了目标函数在积分域上的收敛性，解决了函数不严格可积的问题.

KDE-MEM方法需求解的问题是约束非线性最优化问题，这可以选择序列二次规划法 (SQP)[11]，其优化计算过程可以通过Matlab优化工具箱编程实现，计算中有两点需要注意:

1) 为了用数值方法计算出 (8) 式的积分，必须设置所求积分函数的上下界，一般可由样本分散范围确定;

2) 理论上核函数个数 n 取值越大，密度函数逼近效果越好. 但同时 n 的取值越大，算法复杂度也越高.

4 数值算例

为了验证密度核估计最大熵方法的有效性，以下以两种典型分布为总体进行抽样，使用密度核估计最大熵方法进行分布估计，并将验证结果与经典最大熵方法进行对比分析.

4.1 对数正态分布

对数正态分布的概率密度函数为

$$p_{LN}(x) = \frac{1}{\sqrt{2\pi}\sigma x}\,\mathrm{e}^{-\frac{(\ln x-\mu)^2}{2\sigma^2}},$$

其中 $x>0$, $-\infty<\mu<\infty$, $\sigma>0$.

取 $\mu=0.5$, $\sigma=0.25$，通过计算机从对数正态分布 $LN(0.5,0.25)$ 中抽样，样本容量为200，具体数据从略，根据这些数据计算其前4阶样本原点矩为 $M_1=1.6975$, $M_2=3.0686$, $M_3=5.8923$, $M_4=11.999$.

使用密度核估计最大熵方法方法，选定正态窗函数 $K(u)=\frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2}u^2}$，则总体密度函数的核估计为

$$p_n(x) = \sum_{i=1}^{n} \alpha_i\Big[\frac{1}{\sqrt{2\pi}\sigma_i}\,\mathrm{e}^{-\frac{(\ln x-\mu_i)^2}{2\sigma_i^2}}\Big],$$

为了保证在优化过程中，$1>\alpha_i>0$, $i=1,2,\cdots,n$，且 $\sum\limits_{i=1}^{n}\alpha_i=1$，可以作如下变换

$$\alpha_i = \frac{\exp\beta_i}{\sum\limits_{i=1}^{n}\exp\beta_i},$$

其中 $\beta_i\in R$，这样既满足了 α_i 的条件，也使得优化参数的取值空间变为实数空间，更利于计

算. 所以总体密度函数$p_{LN}(x)$的核估计变换为

$$p_n(x)=\sum_{i=1}^{n}\frac{\exp\beta_i}{\sum\limits_{i=1}^{n}\exp\beta_i}\left[\frac{1}{\sqrt{2\pi\sigma_i}}\,\mathrm{e}^{-\frac{(\ln x-\mu_i)^2}{2\sigma_i^2}}\right].$$

取$n=3$，则$p_n(x)$共包含9个参数，积分区间选取$[0,4]$，其各参数优化的结果见表1，残差平方和$R=3.727\times10^{-12}$. 经典最大熵方法的参数优化结果见表2，残差平方和$R=8.4217\times10^{-8}$.

表1： 对数正态分布下密度核估计最大熵方法参数优化结果

参数	μ_1	σ_1	β_1	μ_2	σ_2	β_2	μ_3	σ_3	β_3
优化结果	2.1304	0.4331	3.7402	1.5519	0.3227	4.7991	2.2919	0.1084	-6.2164

表2： 对数正态分布下经典最大熵方法参数优化结果

参数	λ_0	λ_1	λ_2	λ_3	λ_4
优化结果	-7.9023	6.1686	2.7374	-3.0452	0.5351

根据这些参数值，可以给出密度核估计最大熵方法得到的密度函数与$LN(0.5,0.25)$理论分布的图像，并与经典最大熵密度函数进行对比，见图2. 通过观察图像可以看出，经典最大熵方法的计算结果在右侧拖尾处出现不可积情况. 而最大熵密度核估计分布与理论分布更为接近，且满足概率密度函数基本性质.

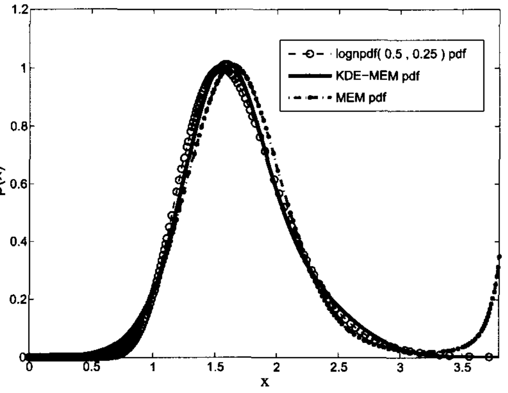


图2： 对数正态密度函数及其估计函数图比较

4.2 Weibull分布

Weibull分布的概率密度函数为$p_W(x)=ba^{-b}x^{b-1}\mathrm{e}^{-(\frac{x}{a})^b}$，其中$x>0$, $a>0$, $b>0$.

取$a=3$, $b=6$，通过计算机从分布Weibull$(3,6)$中抽样，样本容量为200，根据这些数据计算其前4阶样本原点矩为$M_1=2.7830$, $M_2=8.0352$, $M_3=23.9185$, $M_4=73.0769$. 密

度核估计最大熵方法仍选定正态窗函数，取$n = 3$，积分区间选取$[0,5]$，其各参数优化的结果见表3，残差平方和$R = 6.352 \times 10^{-10}$．经典最大熵方法的参数优化结果见表4，残差平方和$R = 2.45 \times 10^{-7}$．

同样可以给出密度核估计最大熵方法得到的密度函数与Weibull$(3,6)$理论分布的图像，并与经典最大熵密度函数进行对比，具体见图3．通过观察图像可以看出，密度核估计最大熵方法分布与理论分布更为接近．

表3: Weibull分布下密度核估计最大熵方法参数优化结果

参数	μ_1	σ_1	β_1	μ_2	σ_2	β_2	μ_3	σ_3	β_3
优化结果	3.0898	0.3687	3.5927	2.5988	0.5360	4.0963	44.5937	60.1679	-15.5288

表4: Weibull分布下经典最大熵方法参数优化结果

参数	λ_0	λ_1	λ_2	λ_3	λ_4
优化结果	-3.3580	7.8318	0.0408	-2.5650	-2.2412

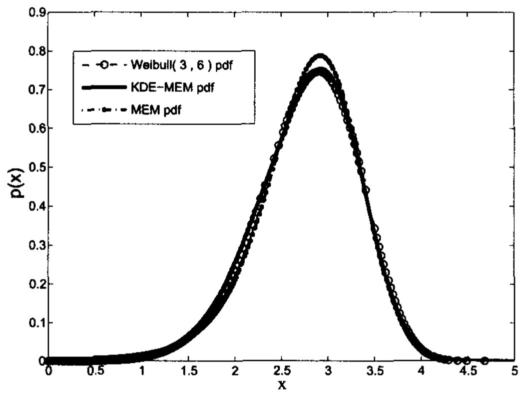


图3: Weibull密度函数及其估计函数图

5 实例

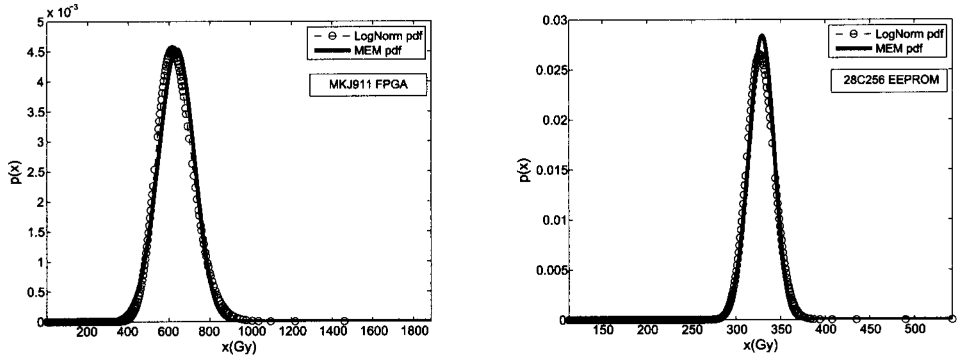
器件MKJ911 (FPGA)和28C256 (E^2PROM)在受到γ辐射时失效样本容量分别为6和5．实验结果表明，这两种器件的失效规律一般遵循对数正态分布，根据失效样本不难估计出对数正态失效分布的参数μ和σ，见表5.

表5: 两种器件失效分布的参数估计

器件	μ	σ^2
MKJ911	6.443	0.01997
28C256	5.792	0.002115

同时根据失效样本，使用本文方法可以得到器件的KDE-MEM失效分布估计．将两种失效分布进行对比，如图4所示．

观察图4可以发现，在不加入任何主观信息的情况下，器件的最大熵分布估计基本能够符合其失效规律．基于失效分布可以估计其失效阈值及方差，将两种失效分布下估计得到的失效阈值及方差进行比较，结果见表6．



(a) MKJ911 (FPGA)　　　　(b) 28C256 (E2PROM)

图 4:　两种器件的对数正态及KDE-MEM失效分布

表6:　两种器件的失效阈值及方差估计

器件	失效阈值/Gy (Si)		方差/Gy (Si)	
	最大熵失效分布	对数正态失效分布	最大熵失效分布	对数正态失效分布
MKJ911	633.147	633.333	83.373	89.592
28C256	327.767	327.800	14.230	15.353

6　讨论与分析

上节中的算例验证了密度核估计最大熵方法的有效性．我们可将两种方法对比分析的结果总结如下：

1)　核估计 $p_n(x)$ 本身满足密度函数的基本性质，优化结果不存在不可积的情况，不需要进行截断就可以直接使用；

2)　在给定样本矩情况下，仅使用3个核函数的核估计，优化结果已经具有较高的精度．可以根据需要，通过增加核函数的数量来进一步提高计算精度．

本文通过改进经典最大熵方法，得到了满足密度函数基本性质的计算结果，使最大熵方法的适用性得到了进一步提高．

参考文献:

[1] Jaynes E T. Information theory and statistical mechanics[J]. The Physical Review, 1957, 106(4): 620-630

[2] 吴乃龙，袁素云. 最大熵方法[M]. 湖南：湖南科学技术出版社，1991
Wu N L, Yuan S Y. Maximum Entropy Method[M]. Hunan: Hunan Science and Technology Press, 1991

[3] 林洪桦. 现代测量误差分析及数据处理[J]. 计量技术，1997, 2: 39-44
Lin H H. Modern analysis of measure error and data processing[J]. Measurement Technique, 1997, 2: 39-44

[4] 陈希孺，等. 非参数统计方法[M]. 上海：上海科学技术出版社，1989: 283-296
Chen X R, *et al.* Nonparametric Statistical Method[M]. Shanghai: Shanghai Science and Technology Press, 1989: 283-296

[5] 詹昊克，姜礼平，苑秉成. 最大熵先验下成败型产品成功率的鉴定试验方案[J]. 工程数学学报，2005, 22(4): 653-658
Zhan H K, Jiang L P, Yuan B C. Reliability qualification test in binomial case based on maximum-entropy priors[J]. Chinese Journal of Engineering Mathematics, 2005, 22(4): 653-658

[6] 王中宇. 测量不确定度的非统计理论[M]. 北京：国防工业出版社，2000
Wang Z Y. Non-statistical Theory in Measurement Uncertainty[M]. Beijing: National Defence Industry Press, 2000

[7] 朱坚民，等. 基于最大熵方法的测量结果估计及测量不确定度评定[J]. 电测与仪表，2005, 42(8): 5-8
Zhu J M, *et al.* Study on evaluation of measurement result and uncertainty based on maximum entropy method[J]. Electrical Measurement and Instrumentation, 2005, 42(8): 5-8

[8] Parzen E. On estimation of a probability density function and model[J]. Annals of Mathematical Statistics, 1962, 33(7): 316-327

[9] 边肇祺，张学工. 模式识别[M]. 北京：清华大学出版社，2000
Bian Z Q, Zhang X G. Pattern Recognition[M]. Beijing: Tsinghua University Press, 2000

[10] 郭照庄，等. 变窗宽密度核估计的构造及均方相合性[J]. 佳木斯大学学报(自然科学版)，2006, 24(3): 423-425
Guo Z Z, *et al.* Structure and consistence of the variable bandwidth for kernel density estimation[J]. Journal of Jiamusi University (Natural Science Edition), 2006, 24(3): 423-425

[11] 简金宝，罗慕华. 一般约束最优化超线性与二次收敛的 SQP 拟可行方法[J]. 工程数学学报，2004, 21(4): 526-530
Jian J B, Luo M H. An SQP quasi-feasible algorithm with superlinear and quadratic convergence for general constrained optimization[J]. Chinese Journal of Engineering Mathematics, 2004, 21(4): 526-530

A New Maximum Entropy Method Based on Kernel Density Estimate

LIU Yu[1], HAN Feng[1], WANG Yu-heng[1], QIAO Deng-jiang[1,2], WANG Jian-guo[1,3]

(1- Northwest Institute of Nuclear Technology, Xi'an 710024;
2- East China Normal University, Shanghai 200062; 3- Xi'an Jiaotong University, Xi'an 710049)

Abstract: For the measurement data with unknown probability distribution, it has been proved that the probability distribution determined by the maximum entropy method (MEM) is a reasonable distribution with the least assumption. But the result of MEM is not always consistent with the properties of a probability density, the resulting truncation and normalization of the result would reduce the accuracy of MEM. A new method based on the maximum entropy principle which combines with kernel density estimate of a probability density is proposed. The estimate given by the new method is totally consistent with the properties of a probability density. The validity and accuracy of the new method are illustrated through examples.

Keywords: measurement data; maximum entropy method; kernel density estimate; distribution estimate

Received: 24 Sep 2009. **Accepted:** 01 Dec 2010.